DOCUMENT RESUME

ED 094 697 IR 000 878

AUTHOR Smith, Michael; Rivoire, Helena
TITLE Shelf-List Conversion.
INSTITUTION Bucknell Univ., Lewisburg, Pa. Ellen Clarke Bertrand Library.; Bucknell Univ., Lewisburg, Pa. Freas-Rooke Computer Center.
PUB DATE 1 Feb 72
NOTE 49p.; For related document see IR 000 879

EDRS PRICE MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS *Cataloging; *Computer Programs; Electronic Data Processing; Input Output; Libraries; *Library Automation; Library Equipment; Library Programs; *Library Technical Processes; *On Line Systems; University Libraries
IDENTIFIERS Bucknell University; Machine Readable Format; Shelf List Conversion

ABSTRACT

The conversion of the shelf-lists of the University Library of Bucknell University to machine-readable form is described in this report. The selection of specific content, machine considerations, and use cost tradeoffs are discussed in the first sections. The actual procedures used in the entering of information into the system, together with the configuration of the computers used are shown to provide a low cost, dependable method for the conversion process. In the first appendix computer facilities and support software are described, while the second contains the conversion manual that was used to instruct students in the operation of the on-line information system used in the conversion. (WDR)

ED 094697

SHELF-LIST CONVERSION

A Joint Project of the:

Freas-Rooke Computer Center

and the

Ellen Clarke Bertrand Library

Bucknell University

Lewisburg, Pa. 17837

February 1, 1972

IR 000 878

Report Prepared by:

Michael Smith, Systems Analyst
Helena Rivoire, Chief Technical Services

Approved by:

E. F. Staiano, Director

G. M. Jenks, Librarian

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TABLE OF CONTENTS

		Page
I.	Introduction	1
II.	Machine-Readable Record Content	2
III.	Conversion Procedures	6
IV.	Summary of Costs	10
	Appendix A: Computer Facilities and Support Software	17
	Appendix B: Shelf-List Conversion Manual	33

Table No.	Description	
1	File Structure	12
2	Conversion Conventions	12
3	Development Expenditures	13
4	Sigma-7 Computer Equipment Description	22

Figure No.	Description	
1	Sample Listing of On-line Conversion Format	14
2	Monthly List of New Books	15
3	Sample Bibliography on Nuclear Science	16
4	Sigma-7 Configuration	21
5	DATASCRIBE Hardware	25

I. Introduction

During the academic year 1969-70 a series of discussions between the University Librarian and the Director of the Computer Center led to the decision that a modest effort should be undertaken to begin the automation of procedures in the Ellen Clarke Bertrand Library. After considerable study of similar projects at other institutions, it was decided that the initial project to be undertaken would be the conversion of the library's shelf-list into machine-readable form. Although certain library procedures could be improved via computer processing without a shelf-list in machine-readable form, it was determined that in the long run the most significant impact upon services offered to users of the library would be made in areas that require a machine-readable shelf-list.

In July of 1970 work began on the conversion of the shelf-list (approximately 190,000 titles; 300,000 volumes) to machine-readable form. The project is scheduled for completion in June of 1972. The purpose of this report is to describe the major elements of that conversion process so that other institutions undertaking a similar project may benefit by the experience gained during this particular effort.

II. Machine-Readable Record Content

Having decided to proceed with the conversion of the shelf-list into machine-readable form, the first and foremost decision to be made was to determine which items on the shelf-list card should be entered into the computer. Since the creation of this record was the principal objective of the project, it is worthwhile to explore the many items considered before the final format for the record was established.

Complete conversion of all information contained on a typical Library of Congress card would, of course, allow for maximum manipulation of the data in the record. However, such a plan presented many difficulties. While recent shelf-list cards in the Bucknell Library are either Library of Congress cards or cards in Library of Congress format, most of the cards over 20 years old contain little more than call number, author, title, and sometimes publication date. Updating the information on the old shelf-list cards prior to conversion was not economically feasible. Moreover, even with an updated shelf-list, the conversion of all information on a complete shelf-list card would have more than tripled the cost of conversion. Thus, it was immediately realized that the information to be converted would be some subset of the total information contained on the shelf-list card.

In the attempt to establish the content of this subset record, two major areas had to be considered. The first area was concerned with the eventual use of the machine record and the items necessary to support those functions. Within this area the following elements were considered: 1) the use of the record for the charge/discharge of books (on-line circulation) 2) the production of new book lists (monthly) 3) the gener-

ation of book lists by Library of Congress classification 4) bibliographic searching both for internal and external use.

With the exception of item 4 all of the above could be satisfied if the record included the Library of Congress number, some form of the title, some form of the author, and the publication date. In order to satisfy completely the bibliographic searching requirement, it would be necessary to specify more precisely the kind of searching to be done. After a brief review of projects at other institutions, it was determined that to undertake an intensive study of bibliographic searching at this time would seriously impede progress on the conversion project. To avoid the reentry of data at some later date, and to permit exploration in the area of bibliographic searching, it was decided that the record should be complete enough to permit automatic matching of records from the MARC tapes. Thus, via machine processing, the record could be expanded at some later time to include all the information contained in the MARC record. (At the present time the Bucknell Library is, in fact, actively engaged in developing the cooperative use of MARC tapes with two other university libraries).

As a result of these deliberations, agreement was reached that complete call number, author, title, and publication date would constitute the minimum data for each record. (In this respect the intended course of action was the same taken by Widener Library at Harvard in the conversion of its shelf-list.[1]) As is quite evident, the goal was shelf-list conversion

[1]De Gennaro, R. "Harvard University's Widener shelf-list conversion and publication program", College and research libraries, v.31, Sept. 1970 pp. 318-331.

simple enough to be completed economically and rapidly, but still containing enough information to be useful.

Having decided on the items to be contained in the machine-record, the second area to be studied was the method of conversion. The availability of two computers on campus, both of which were capable of on-line operation, permitted the consideration of two methods of conversion. The first method, similar to that used at Harvard, would be via keypunching cards, and the second, the use of the on-line facilities of the two computers. Although cost studies on shelf-list conversion projects at other institutions indicated that on-line conversion could be more costly than paper tape or punched cards, an evaluation of the facilities and manpower availability led to the conclusion that on-line conversion would be economically feasible at Bucknell. The ease in editing offered by on-line facilities made possible the use of relatively untrained personnel and students available on a work-study program. A not inconsiderable factor in choosing on-line conversion, its economic feasibility having been determined, was that it offered greater scope for the development of sophisticated error detection programs and other related software.

Although the selection of the on-line entry process offered many advantages, it was deemed necessary to restrain the file size so as not to tax the capacity of the on-line storage of the main computer. This reduction was accomplished by limiting the author entry to the author's last name plus initials only of first names; however, corporate entries were entered in entirety with no truncation. Initial articles in all languages, were omitted both to reduce file size and simplify sorting by title. All titles converted since January 1972 have had the Library of Congress card

number and the international standard book number added to the record whenever such information was available. A complete representation of the items included in the record is displayed in Table 1. All entries in the shelflist, non-circulating reference books as well as special collections, were converted in order to have machine access to the entire library collection.

III. Conversion Procedures

With the establishment of the record content and the decision to carry out the conversion process in an on-line mode, it was now possible to work out the details of the conversion procedures and to begin development of the required software. A more detailed discussion of the available computer systems and the software developed for this project is included in Appendix A. It is recommended that the reader scan the first few pages of Appendix A, as it will help him in understanding the material in this section.

The conversion process proceeded down two parallel paths. The first of these paths was concerned with the conversion of the information on all new books received. This process will continue *ad infinitum* or at least so long as the library is receiving new books. The second path was concerned with the conversion of the items currently in the library. The completion of this segment of the project is scheduled for June 1972. The techniques and programs used in both of these efforts were identical.

From Appendix A the reader will find that the conversion process could be carried out at any of the numerous time-sharing terminals on campus via either the Sigma-7 or the PDP-8. Using the text edit system available on the Sigma-7, or a special program (DATASCRIBE) on the PDP-8, the operator would transcribe the information from the shelf-list card directly to the teletype keyboard.

At the beginning of the project in July 1971, one full-time clerical assistant was employed who worked directly from the shelf-list and keyed in the necessary information on-line either via a CRT or a teletype terminal. She was assisted in this process by students, whose numbers and hours varied depending upon the time of the year and the money at hand. Students

were typically given one hour's instruction by the supervisor. The conversion process was always carried out directly from the shelf-list card; at no time were coding sheets used.

Very early in the project it was realized that there would be a fairly large turn-over in personnel carrying out the transcription process. This was mainly due to the fact that student assistants were to be the primary transcribers and the turn-over of student personnel is quite large. Thus, both simplicity of operation and training were important elements to be considered. A reference manual was prepared for the transcription personnel and is included in Appendix B of this report. The procedures illustrated in that manual, and discussed in this report, constitute an important aspect of the shelf-list conversion project. Careful design of the procedures and the software minimized training requirements and led to a low error rate in the entered data. The error rate, in fact, was so low that one might point to this as being the most innovative part of the project.

A typical day would consist of one or more transcribers creating files either on the PDP-8 or the Sigma-7 that would later be dumped to magnetic tape. The process of copying onto magnetic tape at the end of each day was instituted so as to minimize work lost in the case of computer crashes.

The initial input files were then processed at various times during the week through an error detection program (BOOKCHEK). The output from this program and the editing facilities available on the Sigma-7 permitted the transcriber to go back rapidly and correct those errors detected by the program. When all errors that could be detected automatically had been corrected, a listing of the file was prepared on the computer and the final

human editing process was begun. An example of the output generated by the computer at this stage is included in Figure 1. (Table 2 will serve to illustrate the conventions established as part of the conversion procedure.) The final editing process was carried out by a professional member of the library staff.

When the file was judged fully accurate, punched cards for all circulating books were produced by computer program (CPUNCH).

Conversion was begun with the Library of Congress classification in which books in the Bertrand Library circulate most frequently - "P" - and proceeded through the classification schedule, not from A-Z but in the order of highest circulation. Thus, the most frequently used classifications would be ready first for trial runs and testing of programs for the on-line circulation system. Accordingly, by January 1972 about 90% of the shelf-list had been converted as well as all new titles received during the year. By June 1972 the entire shelf-list will have been converted.

The machine-readable shelf-list has already served purposes other than the circulation of books. A monthly list of new books has been produced each month since October 1970 (Cf. Figure 2). Numerous printouts of the library's holdings in various fields have also been computer generated (BOOKLIST) upon request. For example, the "U" classification listing in booklist form was made for the Military Science program. Further, bibliographies within certain fields have also been produced on request by computer (DKARDS) on 3x5 cards. For example, a bibliography of approxmately 1,000 titles on nuclear science was printed on filing cards by utilizing portions of the shelf-list selected from various classifications relating to nuclear science (Cf. Figure 3).

As the automation of other library procedures is developed, the shelf-list will continue to grow in usefulness. One of the Acquisitions Department routines, for example, will be the checking of incoming order requests by computer against the machine-readable shelf-list to detect potential duplicates. In addition, the machine-intelligible shelf-list will form the base for future on-line searches from remote terminals. Shelf-list data, which can be manipulated in various ways - by author, by title, by keyword, by publication date, as well as by classification number - offer a new and flexible approach to traditional bibliographic information accession.

IV. Summary of Costs

In any project of this nature it is worthwhile to account carefully for all expenditures so that one can evaluate the net worth of the project. A summary of all costs and income is included as Table 3 at the end of this section of the report.

All costs, with the exception of overhead, were obtained from carefully kept records throughout the duration of the project. The overhead was computed at 53% of salaries which has been established as the standard university overhead rate on Federal grants and contracts.

In total, shelf-list information was produced in machine-readable form for more than 190,000 titles. This number when related to the total cost of the project suggests a conversion cost of approximately 42¢ per title (27¢ per volume). The cost per entry for adding new books to the file should be significantly lower than 42¢ per title since only operational costs are involved.

Since the conversion of the shelf-list was a means to an end rather than an end in itself, it is difficult to evaluate the cost effectiveness of this project until such time as related library procedures have been automated. The true value of this project therefore, will not be apparent for several years.

One can, however, compare the cost per title with that incurred by other institutions in similar projects. A review of a large number of such projects indicates that the 42¢ per title cost is quite reasonable. A direct comparison is not possible since each institution selected a somewhat different record content. More significantly, unlike Bucknell, most libraries have not included the cost of actually placing the com-

puter-produced tab cards in the books. In addition, other institutions frequently have not clearly indicated the pro-rated time of the permanent library staff involved in the conversion project or have not considered the entering of new book information (normally a cataloging function) within the scope of the reported conversion expenditures. In this important regard, the Bucknell shelf-list conversion expenditure summary is thoroughly comprehensive.

FILE STRUCTURE

The record for every title in the shelf-list contains, in variable fields:

1) Complete call number.
2) Author's last name followed by initials of first names. Corporate authors are entered completely and without truncation.
3) Complete title (but not sub-title). Initial articles, in all languages, are omitted.
4) Date of publication - latest date shown on card is the date used in cases of variations of copyright and imprint date.
5) Volume number, copy number, or location. Location (Reference, Alumni Shelf, etc.), always preceded by a comma, is the last bibliographic item in the record. Books in special locations do not circulate as a rule. Nonetheless, these titles have been converted to machine-readable form so that the shelf-list can be used for functions other than the circulating process.
6) The Library of Congress card number and the ISBN, if present, constitute the last elements of the record. They are preceded directly by an "=" sign and separated from each other by a comma.

Table 1.

CONVERSION CONVENTIONS

The various fields are delimited by a "+" sign and the bibliographic data on the record terminates with an "=" sign. These signs were chosen somewhat arbitrarily as being symbols not likely to occur in titles. The "<" was used as a sort of ditto mark to indicate a repetition of fields from the previous record when multiple volumes or copies are being keyed in. The various fields are identified by position, i.e., call number, 1st position; author, 2d position; etc. (Cf. Figure 1).

Table 2.

Shelf-List Conversion
Developmental Expenditures

I.	Hardware	
	Rental of 2 terminals (20 mo.)	$1,800
	Special tab cards (80 col.)	600
	Subtotal:	$2,400
II.	Sigma-7 Services	
	Total services used (June 1970 - January 1972)	$37,490
	Total services used (February 1972 - June 1972)	4,200
	Subtotal:	$41,690
III.	Personnel Required	
	Chief-technical services (¼ yr.)	$3,300
	Catalog librarian (1-1/3 man-yrs.)	10,050
	Graduate assistant (½ time for 1 yr.)	2,000
	Undergraduate assistants (2½ man-yrs.)	6,340
	Clerical assistants (½ man-yrs.)	2,330
	Subtotal:	$24,020
IV.	Overhead (@ 53% of Salaries)	$12,730
	Total:	$80,840

Income

NSF COSIP Grant	$15,000
Bucknell Matching Funds	7,500
Freas-Rooke Computer Center	
Sigma-7 Services	41,690
Operational Budget	1,920
Ellen Clarke Bertrand Library	2,000
Bucknell University (Overhead)	12,730
Total:	$80,840

Table 3.

BEST COPY AVAILABLE

Sequence Number for On-Line Editing:

```
1215.000    *V1-64-18210                                  <-- Volume Indicator
1216.000    DS734.7
1217.000    .F3
1218.000    *FAIRBANK, J K
1219.000    *NEW VIEWS OF CHINA'S TRADITION AND MODERNIZATION
1220.000    *1968, REF-68-55793                           <-- LC Card Number
1221.000    DS740.4                                       <-- Location
1222.000    .T79
1223.000    *TUNG, L
1224.000    *CHINA AND THE FOREIGN POWERS
1225.000    *1970-
1226.000    DS775
1227.000    .B513
1228.000    *BIANCO, L
1229.000    *ORIGINS OF THE CHINESE REVOLUTION, 1915-1949
1230.000    *1971-75-150321,0-80+7-0746-4                 <-- ISBN
1231.000    DS777                                         <-- LC Number
1232.000    .T4813                                        <-- LC Number
1233.000    *T'(KHVINSKII, S L                            <-- Author
1234.000    *SUN YAT-SEN                                  <-- Title
1235.000    *1966-                                        <-- Publication date
1236.000    DS777.55
1237.000    .C64
1238.000    *
1239.000    *COMMUNIST CHINA
1240.000    *1955-
1241.000    *1955-58-758
1242.000    <<<X+1967-58-758                              <-- Preceding-field Duplicators
1243.000    <<<X+1968-58-758                              <-- Preceding-field Duplicators
1244.000    DS777.55
1245.000    .S35
1246.000    1968
1247.000    *SCHURMANN, H F
1248.000    *IDEOLOGY AND ORGANIZATION IN COMMUNIST CHINA
1249.000    *1968-
1250.000    DS777.55
1251.000    .V36
1252.000    *VANSLYKE, L P
1253.000    *ENEMIES AND FRIENDS
1254.000    *1967-67-26531
1255.000    DS834.7
1256.000    .H3
```

Figure . Sample Listing of On-Line Conversion Format

Figure 1.

BEST COPY AVAILABLE

PHYSICAL EDUCATION

GV342 .S75 — STREICHER, H
RESHAPING PHYSICAL EDUCATION. 1970.

GV443 .G48 — GILLISH, B C
BASIC MOVEMENT EDUCATION FOR CHILDREN. 1970.

GV1439 .L29A313 — LARSEN, B
LARSEN'S SELECTED GAMES OF CHESS, 1948-69. 1970.

GV1783 .C45 — CHENEY, G
MODERN DANCE. 1969.

GV1799 .H3 — HABERMAN, M
DANCE. 1970.

GV1805 .G777 1970 — FROST, T
CIRCUS LIFE AND CIRCUS CELEBRITIES. 1970.

SOCIAL SCIENCES

H61 .R88 1970B — RYAN, A
PHILOSOPHY OF THE SOCIAL SCIENCES. 1970.

H62 .B42 — BERENSON, C
RESEARCH AND REPORT WRITING FOR BUSINESS AND ECONOMICS. 1971.

H62 .B566 — BLACKMAN, S
INTRODUCTION TO DATA MANAGEMENT IN THE BEHAVIORAL AND SOCIAL SCIENCES. 1971.

H62 .D257 — DAVISSON, W I
INFORMATION PROCESSING. 1970.

H62 .K74 — KRIMERMAN, L I
NATURE AND SCOPE OF SOCIAL SCIENCE. 1969.

H62.5 .A34R4 — RESEARCH IN RURAL AFRICA. 1969.

STATISTICS

HA201 1850 .A42 — U.S. CENSUS OFFICE. 7TH CENSUS, 1850
STATISTICAL VIEW OF THE UNITED STATES. 1854.

BUSINESS AND ECONOMICS

HB74 .M3I57 — INTRILIGATOR, M D
MATHEMATICAL OPTIMIZATION AND ECONOMIC THEORY. 1971.

HB98.2 .F47 — FERGUSON, C E
NEOCLASSICAL THEORY OF PRODUCTION AND DISTRIBUTION. 1969.

HB171.5 .B17 1971 — BAILEY, M J
NATIONAL INCOME AND THE PRICE LEVEL. 1971.

HB199 .D37 — DEAN, R D
SPATIAL ECONOMIC THEORY. 1970.

HB201 .S54 — SHEPHARD, R W
THEORY OF COST AND PRODUCTION FUNCTIONS. 1970.

HB501.5 .T5 — THUROW, L C
INVESTMENT IN HUMAN CAPITAL. 1970.

HB615 .A78 — ARROW, K J
ESSAYS IN THE THEORY OF RISKBEARING. 1971.

HB1360.5 .A78 — ARRIAGA, E E
MORTALITY DECLINE AND ITS DEMOGRAPHIC EFFECTS IN LATIN AMERICA. 1970.

HB3640 .P4W9 — WYON, J B
KHANNA STUDY. 1971.

HB3644.6 .S5S35 — SAW, S H
SINGAPORE POPULATION IN TRANSITION. 1970.

HC60 .A78E — ASHABRANNER, B K
MOMENT IN HISTORY. 1971.

Figure 2.

Figure . Monthly List of New Books

BEST COPY AVAILABLE

HD9698
•A1I6
N860
1969

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS
LABORATORY TRAINING MANUAL ON THE USE OF ISOTOPES AND RADIATION IN ANIMAL RESEARCH. 1969.

R121
•E85
REF

ETTER, L E
GLOSSARY OF WORDS AND PHRASES USED IN RADIOLOGY AND NUCLEAR MEDICINE. 1960.

RA1231
•R2U52

U.S. NATIONAL COMMITTEE ON RADIATION PROTECTION
MAXIMUM PERMISSIBLE AMOUNTS OF RADIOISOTOPES IN THE HUMAN BODY AND MAXIMUM PERMISSIBLE CONCENTRATIONS IN AIR AND WATER. 1953.

Figure 3.

Figure . Sample Bibliography on Nuclear Science

Appendix A

Computer Facilities

&

Support Software

Bucknell University is indeed fortunate to be equipped with as capable and versatile a central computer facility as is maintained by the Freas-Rooke Computer Center. A Xerox Data Systems Sigma-7 serves as this multi-functional computer system (Cf. Figure 4) and is operational in the concurrent modes of time-sharing, remote batch-processing, and local batch-processing. It is one of the fundamental principles of this computer center that the multiplicity of computer requirements demanded by a university of the character of Bucknell should be serviced, to the fullest extent possible, by a central computer system possessing the variety of processors (e.g. COBOL, FORTRAN, Assembly Language, Simulation Languages, File and Information Management Systems, etc.) necessitated by the broad spectrum of university user requirements. But software systems by themselves are of little utility if the computer system is incapable of supporting them concurrently in the modes required, e.g. on-line systems clearly need substantial system capability in a time-sharing capacity. This comprehensive computational capability was thus precisely the support which Bertrand Library felt essential to its automation undertakings, and which has proved admirably equal to the task. A major file-processor-oriented computer dedicated to the automation of library functions was thus obviated, while all essential batch-processing and on-line operations envisaged by the library could be met adequately by the facilities of the Sigma-7. This expectation was, in fact, substantiated in advance through careful analysis of current computer activities and their demands, e.g. the system supports extremely well more than twenty-five terminals located throughout the campus in locations ranging from administrative offices to student dormitories. In sum, therefore, the computer environment

at Bucknell, at the inception of the shelf-list conversion project, was accurately assessed and methodological procedures designed accordingly.

A second and wholly ancillary mini-computer was acquired, however, to serve as a programmable multiplexor/controller for all automation equipment within the library, e.g., teleprinters, video-display units, transaction stations, etc. The device selected, a PDP-8/L, while software was being developed for subsequent projects (for example, on-line circulation), was placed into active use in shelf-list conversion through the creation of a software system, DATASCRIBE, which permitted continued on-line data entry capability as a back-up or alternate path to full utilization of the Sigma-7. While the Sigma-7 was non-functional extremely infrequently for any significant period, the capability of continuity in the on-line conversion approach was of great value both in user convenience, resulting from comparable on-line text entry facilities, as well as in minimizing temporal delays in conversion. Written entirely in assembly language for the PDP-8/L, DATASCRIBE is examined at some length in the following section and should be of substantial interest to any institution considering on-line shelf-list conversion utilizing a very small computer (and very small expenditure).

The additional software routines described in the following document were written for the Sigma-7 in XDS Extended FORTRAN IV and function, except where noted, in both time-sharing and batch-processing environments. These routines are primarily designed for processing of the shelf-list text, error analysis, and file manipulation.

The principal routines are:

1. SORT
2. WEAVE
3. CPRINT, CPUNCH
4. BOOKCHEK
5. BOOKLIST
6. PRINT
7. DUMP, RETRIEVE

Figure . Sigma 7 Configuration.

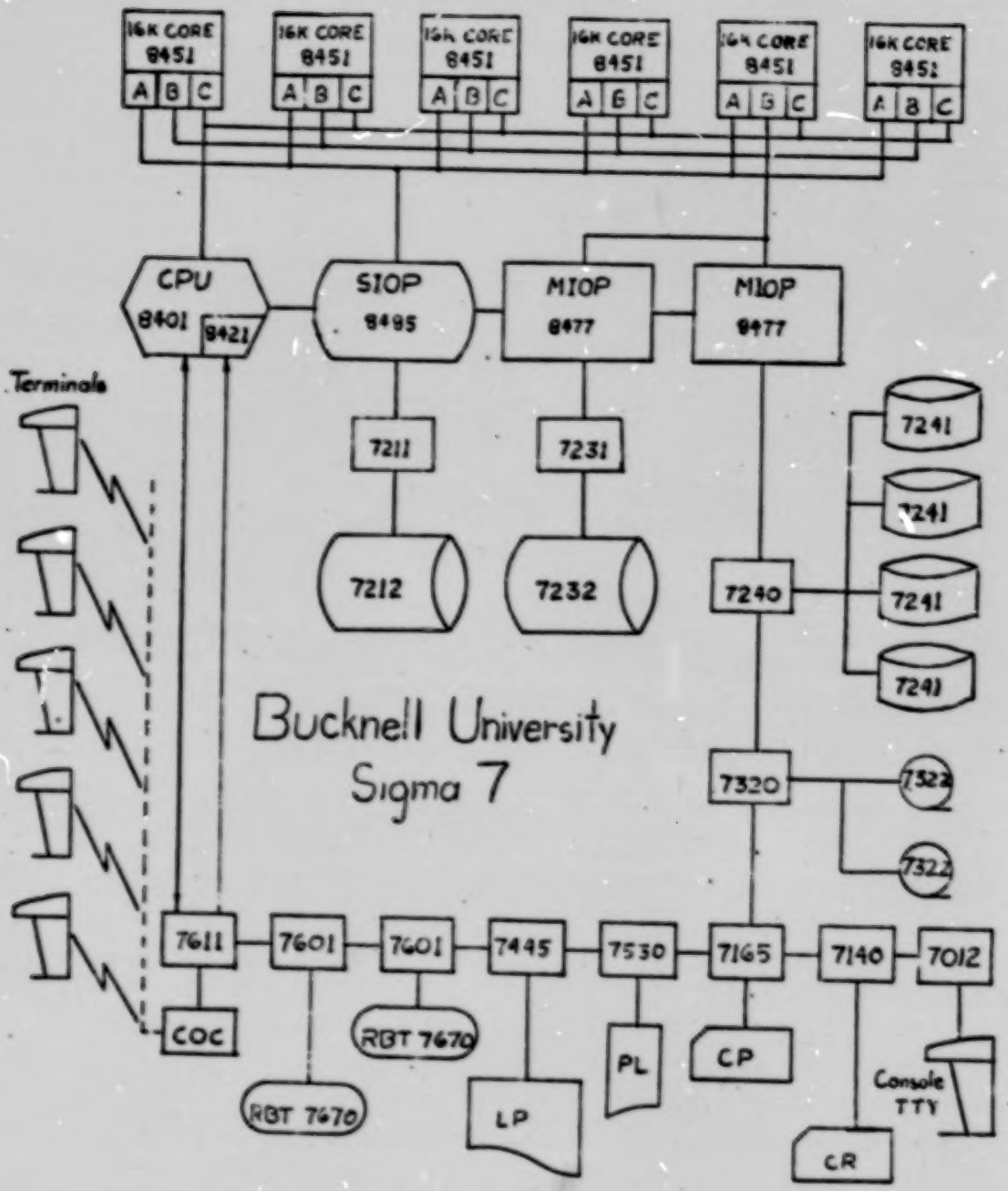

Figure 4.

Sigma-7 Computer Equipment Description

8401	Central Processing Unit
8421	Priority Interrupts
8451	(6) 16K x 32 bit Word Memory Modules (3 way access)
8477	(2) Multiplexor Input/Output Processor
8485	Selector Input/Output Processors
7012	Operator Console
7140	Card Reader & Controller (1500 CPM)
7165	Card Punch & Controller (100 CPM)
7211	RAD Controller (BTM)
7212	BTM Swapping Rad (5.2 mega-byte storage)
7231	RAD Controller
7232	RAD (6.3 mega-byte file storage)
7240	Disc Pack Controller
7241	(4) Removable Disc Pack Drives (100 mega-byte file storage)
7320	Magnetic Tape Drive Controller
7322	(2) Magnetic Tape Drives (9 track)
7444	Line Printer & Controller (1500 lpm)
7530	12" Digital Plotter & Controller
7601	Remote Batch Terminal Controller
7611	Character Oriented Communications Controller (BTM)
7670	Remote Batch Terminals

Table 4.

DATASCRIBE

An indispensable element in any on-line conversion of the shelf-list to machine-intelligible form is adequate support from the computer facility in an on-line mode. If on-line access can be procured with assurances of stability and appropriate operational hours, significant advantages in overall conversion efficiency and time duration can be realized. To achieve these objectives a software system, DATASCRIBE, was developed and constructed in assembly language for the PDP-8/L which could run under the Multi-Task Monitor (MTM/8, an operating system developed at Bucknell) permitting on-line access by the shelf-list conversion personnel at all times of day or night (the latter being a favorite student working period). The shelf-list could then be entered via teletypes and/or CRT's located throughout the campus, the textual data being stored by the PDP-8/L on magnetic cassette. The contents of the cassettes were then transmitted via high-speed telecommunications lines to the Sigma-7, usually once per week. The text files could then be processed further by ancillary Sigma-7 routines as described herewith. Should shelf-list conversion in an on-line mode be considered at other institutions, moreover, the computer system required for such a conversion (viz. a similarly equipped PDP-8/L series system) can be acquired with minimal expenditure.

There are two phases of operation inherent in the data transcribing software system used on the PDP-8/L. The first phase involves the capturing of the data from remote terminals and recording it in a recognizable format on a magnetic tape cassette. In the second phase, the data are read from the cassette and transmitted to the Sigma-7 where respective files are created on disc.

The primary advantages realized with a system of this nature are as follows:

- As many as four communications channels can be dedicated strictly for transcribing work by library personnel for long periods of time.
- There is less chance of losing data due to the occasional hardware and system malfunctions of the Sigma-7. If data should be lost on the disc files, they are easily reconstructed from the cassette files.

Hardware Requirements: (Cf. Figure 5):

PDP-8/L with the following options:

1.) Power Failure Detection and Restart
2.) 60 Hz Real-Time Clock
3.) Memory Extension Unit (3K x 12)
4.) ASR 33 Teletype
5.) DC02 Multiple Communications System with 1 to 4 Datasets (110 to 1200 baud)

Sykes Compu/Corder Model 100.

Functional Description of the Data Collection and Storage Program: (Phase 1)

Multi-user DATASCRIBE (MDS) is a program written to receive data from up to four independent users simultaneously via the DC02 multiple channel asynchronous communications system. These data are then identified as coming from a particular user and recorded on the magnetic

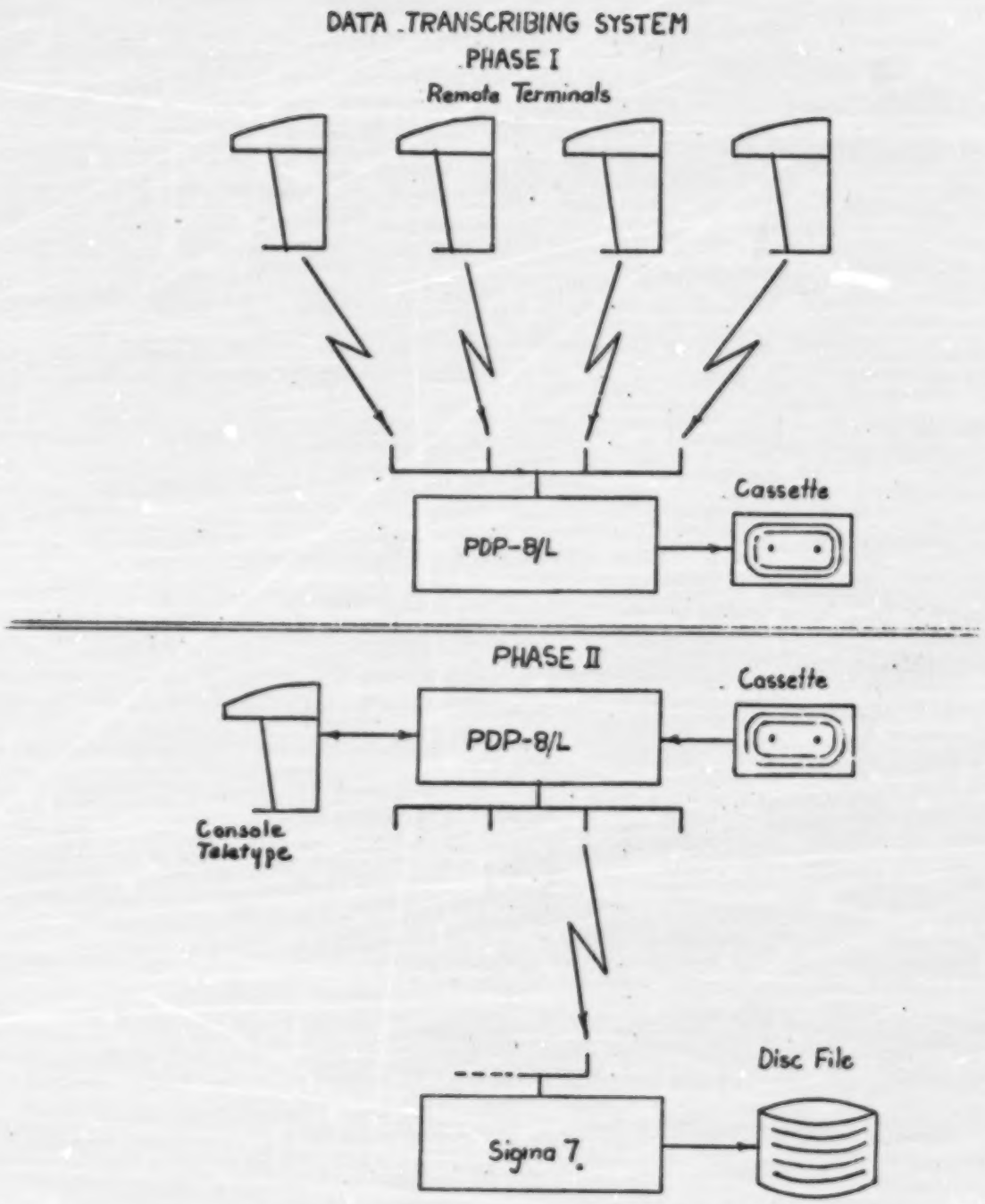

Figure . DATASCRIBE Functional Components.

Figure 5.

tape cassette. Preliminary examination and editing of the characters is effected prior to their being recorded to determine that illegal characters are not accepted and to examine input for special control characters which cause MDS to initiate special actions. The system is designed to simulate the operation of the text editing subsystem operating on the Sigma-7 (BTM) so that there is minimum confusion for the people doing the transcribing when they are forced periodically to alternate between work on the PDP-8 and the Sigma-7. For example, the following conventions used on the Sigma-7 were also adopted for use with MDS:

ESC and ALT MODE are treated as equivalent

<ESC><RUBOUT> echoes <←>

<ESC><X> echoes <RET><LF>

<RET> echoes <RET><LF><-><SPACE>

The remaining operating procedures required of a user are very basic. After dialing into the DCO2 communications system, the user is recognized by the computer when the program receives the first character from the user. It responds with READY, and from this point the text received will be recorded (as edited by the user) on the cassette as data. When the user wishes to sign off and close his file on the cassette he issues a simple terminating command to the PDP-8.

Because the MDS program is capable of concurrently servicing up to 4 users, any number of which may require servicing at or near the same time, all of those routines that are commonly used by all four users are written in a re-entrant code. Each user has his own set of variables and input and output buffers reserved to himself, thus saving on memory storage requirements and allowing greater versatility. Whenever a user requires service

(i.E. he sends or receives a character from DC02), two distinct levels of logic are called upon. First, there is a set of small, fast routines which establish the set of variables to be used. After this has been accomplished, execution of the common routines, which actually service the user, are initiated.

Records recorded on the Compu/Corder are written at prerecorded, directly addressable tape addresses. Each record contains 126 characters of data, the first of which is reserved as an identifier of the user. As each record is written in the C/C, the tape address (TA) at which it was written and the identifier of the user (ID) is also recorded in the upper 4K words of core memory in a special table. After all users have signed off, this table is written at the beginning of the cassette.

MDS also has provisions for concurrent and continued checking of potential hardware malfunctions and power failure (i.e., special purpose diagnostic routines). Depending upon what routines were being executed and what devices were in operation at the time, appropriate action is taken to record the symptoms of the error. An attempt is made to recover with a minimum loss of time and data, and if recovery is found to be impossible, notification is given to the operator.

Functional Description of the Transmitter Program: (Phase 2)

The primary function of this program is to read magnetic tape cassettes containing files built by the previously described MDS program and transmit them to the Sigma-7 computer via the DC02 communications system.

Operation of this program requires a basic knowledge of the use of the time-sharing operating system (BTM) on the Sigma-7. After the cassette has been mounted, the program loaded into the PDP-8/L memory, and execution started, the operator is in communication with the Sigma-7 via

the console teletypewriter. At this time he must initiate execution of the Sigma-7 data reception routine. The program operating in the PDP-8/L is designed to retrieve from the cassette and transmit to the Sigma-7 all those records relating to a single user before proceeding to the next user (if more than one). Even though there may be records relating of up to four users, the entire file for any one user is completely retrieved from the cassette and transmitted to the Sigma-7 before the records of the remaining users are processed.

Provisions have also been made in the program to detect hardware errors or power failures during operation and to make suitable attempts at recovery from these error conditions. Facilities are provided, in addition, for monitoring the rate at which the Sigma-7 accepts data transmitted from the PDP-8/L, so the program will only transmit data as fast as the Sigma-7 is capable of receiving and processing it. This facility insures that data communications buffer allocations in the Sigma-7 will not be exceeded, thereby avoiding the loss of data during transmission.

SORT

SORT arranges a file of book information modules in call number ascending order (Library of Congress regimen). The program uses four sequential files, all generally disc-based. The SORT routine, a polyphase replacement type, utilizes main memory as fully as possible during the sorting process, resorting to sequential disc file data manipulation only upon exhausting the allocation of core memory. SORT functions in both time-sharing and batch-processing modes.

WEAVE

WEAVE, a serial mode, reads book information modules from two independent pre-sorted input files, and by comparing the call number of each module (as well as volume delimiting identifiers if necessary), decides which module has call number precedence. The selected module is then written into the output file and another module is read from the previously accessed input file. This process of reading, comparing and writing continues until both input files have been read completely. It is assured that each input file has been previously sorted so that the book information modules are arranged by call number according to strict LC formal ordering principles. WEAVE operates in both time-sharing and batch-processing modes.

CPRINT, CPUNCH

These two programs are identical except for the output devices used (CPRINT employs the line printer for output; CPUNCH, the card punch). Each program reads a file containing a number of book information modules. (Cf. Program BOOKLIST). As each module is read, the program determines whether this particular volume has any special location in the library, or is identified as other than a phonograph record, an oversize volume, or an Asian Collection volume; if so, the volume entry is ignored. If a card is to be punched, a standard output record such as the following is created:

F3326 .M25 MALLOY, J M BOLIVA: UNCOMPLETED REVOLUTION

Special information, such as volume and/or copy number may appear. If an author's name is longer than 40 characters, the program determines where to terminate the name so that it is as close to 40 characters in length as possible without truncating a word. The title field is ex-

tended as far as the eightieth character of the output record. When a record has been created and formatted, it is either printed or punched, depending on the output device used. CPRINT, which prints the lines, may be used to check on the correctness of the output. CPUNCH does the actual card punching.

BOOKCHEK

BOOKCHEK examines the contents of the one-to-three files containing book information modules in order to indicate possible syntactic errors. Several assumptions were made about the syntax and form of the modules, namely:

1. No line contains more than two consecutive embedded spaces. Moreover, the first two lines of each module contain no embedded blanks.
2. No line contains more than eighty characters.
3. No module contains fewer than five lines nor more than twelve lines.
4. The first line of each module must begin with an alphabetic character (Call Number).
5. The second line of each module must begin with a period.
6. The number of information fields preceded by a plus (+) must be either three or four. If there are four such fields, the fourth must begin with a 'V'.

If any violations of these and other more intrinsically structured assumptions occur, the line number containing the potential syntactic error and a diagnostic message are written on the line printer. At the end of program execution, the total number of lines read and the message "NO MORE ERRORS DETECTED." are printed. BOOKCHEK functions only

in the batch-processing mode.

BOOKLIST

BOOKLIST prepares an output file of pre-processed book information modules. Each module consists of several lines of information, including call number, author, title and date of publication of each volume incorporated in the input file. In addition, other information that may be necessary to identify the book uniquely, such as volume and/or copy numbers, as well as special location within the library, such as the reference room, may be included.

The program reads the input file three times. During the first pass, the routine selects all those book information modules which correspond to books classified as reference volumes. The second time, it selects all other modules except those corresponding either to reference volumes or to phonograph records. As each module is selected, the program decides whether a new topic heading is necessary. If so, the topic heading is selected from a 'heading' file and written into the output file. Each module is then arranged in the following standard form:

```
Columns:    1       10  12                                    61

Call no.:   HD7049  INTERNATIONAL RESEARCH ASSOCIATES          <-- Author
            .I53        NEW FAR EAST:  SEVEN NATIONS OF ASIA  <-- Title
Lib. Loc.:  REF.        1966.                                 <-- Date of Pub.
```

After each module is arranged in this form, it is written into the output file.

As the program is running, various messages may be transmitted to the remote terminal. Any lines in the input file which begin with "*", "<", or "/" are so transmitted. If the first character of a module's call number is not a letter, the message "NON-ALPHA. CHAR. IN CALL NO",

the first three characters plus the number of numeric digits in the call number are printed.

Upon completion of the output file generation, the routine submits a batch-processing job which directs the contents of the output file to the line printer. The processed book modules are printed on 11 inch x 15 inch paper in a double column format. Additional character checking is performed prior to printing each module, and each page is terminated by page number line. The printed output, finally, is photo-reduced to achieve the standard book list form (Cf. Figure 2).

PRINT

PRINT is a multi-functional routine which lists the shelf-list file in its internal or unformatted structure. This listing is directed to the line printer only and, accordingly, operates in batch-processing mode exclusively. The user selects, at the time of job submittal, the incremental value by which listed lines may be separated (e.g., 1.000, 0.100, etc.). These listings are then used by conversion personnel for proofreading and editing purposes.

DUMP, RETRIEVE

These two routines perform the daily requirements of saving conversion files on magnetic tape and retrieving selected files as necessitated by subsequent desires for listing or editing these files. DUMP and RETRIEVE jobs are submitted to the batch-processing stream through interactive time-sharing terminals.

Appendix B

Shelf-List Converter's Manual

The following document is a copy of the manual distributed to beginning shelf-list typists which has been found quite adequate, coupled with minor oral instructions, to permit the student to translate the shelf-list card into machine-intelligible format and enter it into the computer system. This straight-forward character of textual entry cannot be valued too highly in evaluating an on-line conversion project; its simplicity and lack of redundancy (e.g., in bypassing the use of encoding sheets for keypunchers) are primary assets in undertaking rapid and economical shelf-list conversion.

Though much detail in the following manual is meaningful only in the context of the Bucknell community and computer system, the reader will undoubtedly grasp the significance of the instructions and the flexibility engendered by the on-line approach to shelf-list conversion.

TELETYPE TERMINAL

EDIT subsystem

Depress ORIG button on lower right.

Dial 1801, 1701. 3291, 3287, 1185, or 1097

When beep is heard, teletype will begin to type:

```
LOGIN:  00610,201, LIBRARY <RET>                    The word LIBRARY will not
                                                    appear on the paper.
!EDIT

*BUILD SL123 <RET>                                  To start a file

1.000********SL123    DATE    YOUR NAME********<RET>
2.000 <RET>                                         Hit return without typing
                                                    to receive a prompt.
*EDIT SL123<RET>

*IN 2,.1 <RET>                                      This command causes the
                                                    computer to insert line 2
                                                    and succeeding lines
                                                    incrementing by .100.

2.000    PR2300 <RET>
2.100    .F56 <RET>
2.200    T969 <RET>
2.300    +FROST  F <RET>
2.400    +WORKS <RET>
2.500    +T969= <RET>
2.600    PR230T <RET>

           etc.

25.600   ******** <RET>                             **** to indicate last line
25.700   <RET>                                      of work before logging
*END <RET>                                          off.  Keep a note of
                                                    last line number.
!BYE
```

Computer statistics follow.

Teletype turns itself off.

You need to type only the underlined parts; the computer types all other characters.

EDIT subsystem

TO CONTINUE A FILE

Follow directions on previous page through:

LOGIN etc.

! EDIT

*EDIT SL123 <RET>

*IN 25.7,.1 <RET>

Line no. following last line of previous input. Be sure to indicate increment when you start a file.

```
    25.700  *********SL123   4 JAN  1971  PM   YOUR NAME *******   <RET>
    25.800  call no.  <RET>
    25.900  call no.  <RET>

            etc.

    50.800  ************  <RET>
    50.900  <RET>

*END  <RET>
!BYE
```

Computer statistics follow.
Teletype shuts off.

TYPE ONLY THE PARTS UNDERLINED

CORRECTIONS

<ESC> <RUBOUT>	erases a letter at a time on the line currently being worked on
<ESC> < X >	erases entire line being worked on and shifts carriage to a new line without adding a new line number

If an error is detected after a line shift, hit the <RET> key on an untyped line to get the prompt * ; then insert the line number to be corrected. Retype.

```
    100.000 PR2300  <RET>
    100.100 .F56    <RET>
    100.200 T969    <RET>
    100.300 +FTOST, F <RET>          If you notice the error in
    100.400 +WORKS  <RET>            line 100.300 at this point,
    100.500 <RET>                    hit the  RET  key.

*IN  100.300 <RET>

    100.300 +FROST, F <RET>

    (Bell)
*IN 100.500 <RET>                    Now you are back on the
                                     line needed for continuation.
    100.500
```

SHELF-LIST CONVERSION

What you are going to do is to put a selected amount of information from shelf-list cards into machine-readable form.

We call one catalog card a RECORD.

From each card or RECORD we are selecting certain elements to be put into the computer - call number, author, title and publication date. Because we want the computer to be able to separate these elements, we put a distinctive mark between the elements. We have been using the + sign to indicate the beginning of a new element within a record. When we finish a record, we put =.

The various elements on the catalog card which we want the computer to recognize and keep separate are called FIELDS. Four fields must always be indicated:

1) call number
2) author — If there is no author, put + and leave the rest of the line blank
3) title
4) date — If there is no date, put + and leave the rest of the line blank.

Additional fields may be indicated to record volume numbers or LOCATION notations - Ref., Spec. Coll., etc.

Information on the various FIELDS will be found on the pages following.

FIELD - CALL NUMBER

PR2300 .F56 1969		Type just as it appears on the card except for LOCATIONS preceding or following the call number
Q PR2300 .F56	PR2300 .F56 + author + title + date, OVER=	Q means that the book is shelved in the OVERSIZED section. Omit Q in call number and add in Location field.
PR2300 .F56 Ref.	PR2300 .F56 + author + title + date, REF=	
R ML100 .S6	ML100 .S6 + author + title + date, REC=	R indicates a phonorecord.
PR2300 .F56 E.P.	PR2300 .F56 + author + title + date, EP=	E.P. indicates that book is in the Enoch Perrine Library in Vaughan Literature Building.
PR2300 .F56 Cage	PR2300 .F56 + author + title + date, SPEC COLL=	Cage is old terminology for Special Collection. Use SPEC COLL in Location field for all books marked Cage.
PR2300 .F56 Alumni Shelf	PR2300 .F56 + author + title + date, ALUM=	

You will also encounter additional locations such as CHEM LAB, ASIA LIB. Enter these locations in the same manner as indicated above.

	FIELD - AUTHOR	
Frost, Francis	+FROST, F	+ to indicate separation between call number and author fields. Type author's last name followed by initials of first names. Omit periods after initials.
American Society of Mechanical Engineers	+AMERICAN SOCIETY OF MECHANICAL ENGINEERS	+ to indicate new field. Type corporate authors in full even though two or three lines may be necessary.
	(no author) Book listed under title	+ (rest of line blank)
De La Mare, Walter Van Doren, Carl De Schweinitz, A	+DELAMARE, W +VANDOREN, C +DESCHWEINTZ, A	Since names with particles preceding them are filed as one word, type these names as one word.
Campbell-Bannerman, H	+CAMPBELL BANNERMAN, H	Double names to be typed as two words without hyphen.
Douglas Home, Alexander Douglas-Home, 14th earl of,	+DOUGLAS HOME, A D H	
Goethe, Johann Wolfgang von	+GOETHE, J W VON	Type out the particle following a name instead of using as an initial.
Vigny, Alfred de	+VIGNY, A DE	
U.S. Library of Congress. Legislative Reference Service	+U.S. LIBRARY OF CONGRESS. LEGISLATIVE REFERENCE SERVICE	Keep punctuation between subdivisions of main entry.

FIELD - TITLE

27 tales	+27 [TWENTY SEVEN] TALES	Spell out, between square brackets, in the language of the title, any numerics that appear at or near the beginning of a title.
1001 nuits	+1001 [MILLE ET UNE] NUITS	
Un coeur simple	+COEUR SIMPLE	Omit definite and indefinite articles in any language at the beginning of a title. Retain the article anywhere within the title.
A Shakespeare companion	+SHAKESPEARE COMPANION	
Die Blechtrommel	+BLECHTROMMEL	
Bio-bibliography of the Kennedy family	+BIOBIBLIOGRAPHY OF THE KENNEDY FAMILY	If the two parts of a hyphenated word do not form two separate words, omit hyphen and type as one word. If the two parts can be read as separate words, omit hyphen and type as two words.
Grain-mills and flour in classical antiquity	+GRAIN MILLS AND FLOUR IN CLASSICAL ANTIQUITY	
China and the U.S.	+CHINA AND THE U S	Omit periods and insert spaces between U S when part of title.

Titles in non-Roman character (Russian, Japanese, etc.) are transliterated at the bottom of the card. Copy this transliteration for the title.

FIELD - DATE

1969,c1968	+1969	Two dates separated by a comma, use latest date.
1968-	+1968-	Open date is a date followed by a dash and nothing or a pencilled date. Use only the first date and dash.
1969-70	+1969-70	Closed dates. Both dates are printed or typed in. Use both dates.
	+	+ and rest of line blank indicates that no date is given.
1969/70-	+1969/70-	Slashed dates.

FIELD - LOCATION

PR2300 .F56 Special Collection	+date, SPEC COLL=	LOCATION is always given as the last field of a record and is preceded by a comma instead of + .
Q PR2300 .F56	+date, OVER=	Q on a line by itself means OVERSIZED.
R PR2300 .F56	+date, REC=	R on a line by itself means phonorecord.
PR2300 .F56 E.P.	+date, EP=	E.P. means Enoch Perrine Library in Vaughan Literature Building.
PR2300 .F56 Ref.	+date, REF=	Book is shelved in the Reference Department.

Volumes and copies are omitted from records having a LOCATION field, except Q (OVER) and R (REC)

```
AE5
.E333
1970
+    (no author)
+ENCYCLOPEDIA AMERICANA
+1970, REF=
```

No mention of volumes or copies when there is a LOCATION field.

```
AE5    Encyclopedia Americana.  New York,
.E333      Americana Corp., 1970.
1970       30 v.  illus.  28 cm.
Ref.
```

FIELD - ADDED VOLUMES OR COPIES

```
PGl   Revue des études slaves.  t.
.R4       Paris, Imprimerie Nationale [etc.]  1964-


289400  v.43
289401  v.44
289402  v.45
289403  v.46
289404  v.47
```

```
PGl
.R4
+
+REVUE DES ETUDES SLAVES
+1964-
+V43=
<<<X+V44=
<<<X+V45=
<<<X+V46=
<<<X+V47=
```

+ and blank line - no author

+1964- use first date only as second date is in pencil. Add the first noted volume to the end of the first record. <<<X+V44= tells the computer to repeat the first three elements of the previous record - call no., author, and title -, to omit the date (X) which is the 4th element, and to add new volumes no. (V44) to call number.

```
AE5
.E333
1970
+    (no author)
+ENCYCLOPEDIA AMERICANA
+1970, REF=
```

```
AE5     Encyclopedia Americana.  New York,
.E333       Americana Corp., 1970.
1970        30 v. illus.  28 cm.
Ref.
```

No mention of volumes or copies when there is a LOCATION field.

FIELD - ADDED VOLUMES OR COPIES (con't.)

For multiple volumes or copies of records or oversize books, the location must be indicated after each volume rather than in the date field alone, e.g.,

```
Call No.
+AUTHOR
+TITLE
+1964
+V1, OVER=
<<<X+V2, OVER=
<<<X+V3, OVER=
```

Any line beginning with a < must end with an equal sign, sometimes requiring that two fields be entered on the same line, e.g.,

```
Call No.
+AUTHOR
+TITLE
+1967
+V1
+C2=
<<<X+V1+C3=
<<<X+V2=
```

When multiple copies are both circulating and/or have been assigned a location, <'s must be used.

```
c.[1]  reference
c.2    circulates
```

```
Call No.              Call No.
+AUTHOR               +AUTHOR
+TITLE                +TITLE
+1970        or       +1970, REF=
+C2=                  <<<+1970+C2=
<<<+1970, REF=
```

The X is omitted since the altered date field must be in the fourth field.

LC (Library of Congress) Number and ISBN (International Standard Book Number)

With each new drawer that is to be entered into the computer, the LC card number will be included. It is found in the lower right corner of the shelf list card and should be entered immediately following the equal sign in the last field of the record. Disregard any alpha prefix or suffix in the number. For multiple copies or volumes, the number is needed only after the first copy or volume. See example below.

PR2322
.R4

Reynolds, Ernest Edwin, 1894–
Sir Thomas More, by E. E. More. [London] Published for the British Council and the National Book League, by Longmans, Green [1965]

83 p. port. 22 cm. (Writers and their work, no. 178)

Bibliography: p. 33-35.

1. More, Sir Thomas, Saint, 1478-1535. (Series: Bibliographical series of supplements to British book news on writers and their work, no. 178)

PR2322.R4 70-17009

942.05'2'0924 [B]

MARC

Library of Congress 70 [7]

PR2322
.R4
+REYNOLDS, E E
+SIR THOMAS MORE
+1965=70-17009

LC (Library of Congress) Number and ISBN (International Standard Book Number)

For the new book lists, the LC card number and the ISBN are to be included. The ISBN is found in the lower left corner of the shelf-list card, and should be entered after the LC card number separated by a comma. See example below.

QA164
.C64

Combinatorial mathematics and its applications; proceedings of a conference held at the Mathematical Institute, Oxford, from 7-10 July, 1969, edited by D. J. A. Welsh. London, New York, Academic Press, 1971.

x, 304 p. 24 cm. B***

Organized by the Mathematical Institute, Oxford.
Includes bibliographical references.

1. Combinatorial analysis—Congresses. I. Welsh, D. J. A., ed. II. Oxford. University. Mathematical Institute.

QA164.C64 511'.6 75-141734
ISBN 0-12-743350-3 MARC
Library of Congress 71 [4]

```
QA164
.C64
+
+COMBINATORIAL MATHEMATICS
AND ITS APPLICATIONS
+1971=75-141734,0-12-743350-3
```

FREAS-ROOKE COMPUTER CENTER
BUCKNELL UNIVERSITY
LEWISBURG, PA. 17837